ChinaXiv合作期刊
图书情报工作
第64卷 第1期 2020年1月

高校图书馆数据流失与保全对策*

■ 李玉海[1,2] 张驰[1] 王文卓[1]
[1] 华中师范大学信息管理学院 武汉 430079 [2] 华中师范大学中国图书馆创新发展研究中心 武汉 430079

摘 要：[目的/意义]调查我国部分双一流高校图书馆数据保全能力以及流失现状，研究和分析数字保存能力成熟度模型在减少数据流失以及数据保全的应用，为国内数据保全系统的建设和发展提供参考。[方法/过程]通过对国内外部分高校图书馆官方网站的数据保全和利用情况进行探究，从数字馆藏资源、数字保存能力与政策、数字保存意识等维度进行分析，并对图书馆运作过程中产生的数据流失现象提出意见。[结果/结论]总结我国部分高校数字保存现状，提出我国高校图书馆提高数字保存能力的必要性与重要意义，并用数字保存能力成熟度模型（DPC-MM）对高校图书馆数据保全提供理论支持。

关键词：高校图书馆 数据流失 保全
分类号：G251
DOI：10.13266/j.issn.0252-3116.2020.01.017

chinaXiv:202304.00374v1

1 引言

随着数字化时代的迅速到来，数字资源的保存和保全已然成为一个艰难的问题。数字资源的管理复杂度高、风险大，各行各业的数据保全面临着成本高、价值密度低、保护政策缺失等重大问题。而图书馆、博物馆、艺术馆等机构作为社会公共组织，在人类数据保全中起到至关重要的作用，在国外高校图书馆数据保全已逐渐成为热点议题。笔者通过分析国内部分高校图书馆官网中数据相关政策以及利用程度时发现，国内高校图书馆在数据保全方面仍存在重大缺陷，在高校图书馆运作过程中出现了数据流失现象，造成科学管理的不完善以及数据无法追溯等相关问题。

近年来，国外学者和机构开始研究如何利用能力成熟度模型（capability maturity model, CMM）来评价数字保存系统，而我国只有极少数学者借鉴相关模型进行研究和探索，其中国内对数字资源保存的不重视占据了相当一部分原因。能力成熟度模型的引入，在一定程度上可解决数据保全系统可信赖指标细化和量化的难题，为数据保全系统迭代与更新提供了理论支持与方向，同样为数据流失的解决方案提供了有效的依据。其中，数字保存能力成熟度模型（digital preservation capability maturity model, DPC-MM）在数字保存领域应用的最广泛且相对成熟，笔者借用DPCMM模型探讨国内高校如何解决数据流失的难题以及提出科学有效的解决方案。另需注意，笔者所提及的保全并非指将图书馆生产以及尚需挖掘的所有数据进行保存，而是指在图书馆由于成本过高等条件限制下，对其可控范围且具有价值的数据进行保全。

2 图书馆数据及其特点

图书馆的数据类型较多，如基于移动端的用户交互数据、基于Web端的网页使用数据、基于图书馆实体的管理数据等。而由于图书馆的特殊性，图书馆的数据又有学术专业性、业务逻辑严谨性、对外服务友好性及可观赏性等多种特性，本文所指图书馆数据主要为已被发现且具有一定价值的数据，不包含尚未挖掘的价值性数据。

2.1 数据及其来源

数据是文字和符号的集合，经逻辑归纳而形成的对客观事务的表示素材，具有分析和加工的意义。一切有记录的、有特殊意义的文字和符号都是数据。

* 本文系中央高校基本科研业务费重大培育项目“智慧图书馆管理系统关键技术与应用研究”（项目编号：CCNU18JCXK04）研究成果之一。
作者简介：李玉海（ORCID：0000-0003-2256-0183），院长，教授，博士生导师，E-mail：yhli@mail.ccnu.edu.cn；张驰（ORCID：0000-0001-5784-7323），本科生；王文卓（ORCID：0000-0002-3702-9814），本科生。
收稿日期：2019-12-02 本文起止页码：139-145 本文责任编辑：徐健

ChinaXiv合作期刊

数据体现在方方面面。例如,手机上的 QQ、微信等无时无刻不在产生数据。换句话说,与人类行为、活动有关的信息都可以被称为数据。在高校图书馆,数据可体现在资源和服务两个方面[1],书籍的编号、电子数据库中的馆藏信息、计算机上及书的文字、图书馆的温度监测系统、图书馆内某一个座位的使用频率等都是具有利用价值的数据。

当今信息化环境下,图书馆的数据来源主要有 3 种渠道:一是实体图书馆产生的生产服务数据,二是图书馆网站的资源利用数据,三是移动终端的用户行为数据。如图 1 所示:

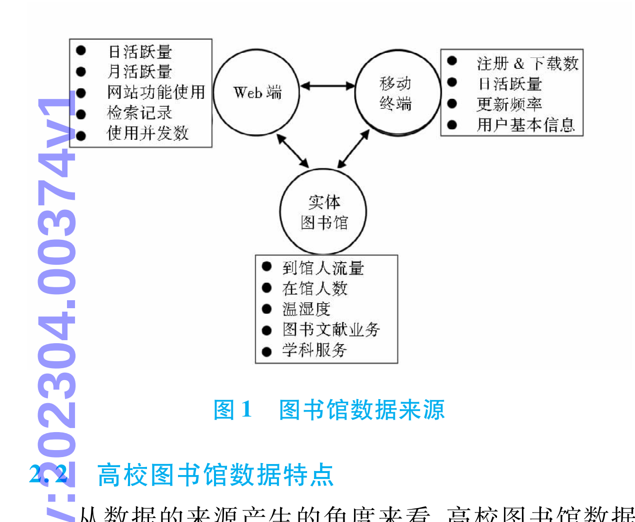


图 1 图书馆数据来源

2.2 高校图书馆数据特点

从数据的来源产生的角度来看,高校图书馆数据具有动态性、季节性和时效规律性。动态性体现在高校图书馆数据随图书馆资源使用情况而变化,例如书籍借阅量、座位使用情况;季节性和时效规律性体现在数据是动态变化的,但其具有一定规律,例如高校考试周图书馆入馆人数明显高于其他时间段。

从存储利用的角度来看,高校图书馆数据具有以下特点:①基于文献数据的学科专业性。高校图书馆馆藏多以学科分类,面向对象为在校大学生,书籍多为权威教材及参考资料。②基于生产管理数据的客观严谨性。高校图书馆数据的管理根据图书检索类型与频次、书籍与个人借阅次数以及借阅时间段进行划分,具有一定客观性,经管理人员分析以调整策略。③基于读者需求与对应书籍数据的良好匹配性。不同对象产生的数据不同,可根据数据集分析得出相应策略。例如大学新生与老生的书籍需求不同,不同专业侧重(理工科、文科)的大学专业书籍馆藏类别和数量不同。④基于数据管理服务的周期性。图书馆所生产的数据具有一定规律性,图书馆可通过其规律调整服务政策。

3 图书馆数据类型

从数据流失现象和保全需要可对图书馆数据进行以下分类:

3.1 按照可观测程度

从可观测程度来看,高校图书馆数据可分为显形数据和隐形数据。显性数据是指公众可浏览的可视数据,如期刊杂志、数据库、书籍上的文字,书籍的馆藏状态,书籍、期刊的编号等;隐形数据是指部分人可视的随机出现的客观数据,如图书馆电脑机房中某台电脑上一天内产生的浏览数据,自习室中某一区域座位一天内的使用人数,某一学科书架上的书籍流动情况等。

3.2 按照业务流程

从业务流程角度来看,高校图书馆数据可分为生产型数据、管理型数据及服务型数据。生产型数据指图书馆中的资源在组织、加工过程中所产生的数据,如纸质图书的条码,电子资源的分类整理,文献的组织、加工等;管理型数据用于文献、机构等的管理,主要针对图书馆馆长、主任及书记等,如图书馆内图书总量、期刊数据库来源、管理经费、图书馆流动人员实时监测数据等;服务型数据主要为读者服务,如索书号、图书馆座位预约编号等。

3.3 按照数据存续时长

根据存在时间长短,高校图书馆数据还可分为临时数据、中转数据和长期数据。根据数据的价值,将数据的保藏时间进行严格管理,有利于资源的存储及调用,增加存储效率。针对不易保存的数据,构建完善的保存体系,加强数据的管理。

3.4 按数据产生对象

从数据产生对象的角度,高校图书馆数据可分为服务请求数据和服务响应数据。服务请求数据是指读者、客户等服务接受者在使用图书馆的过程中通过人工窗口或图书馆内电子设备发出的服务请求数据,最常见的有图书的借阅和电子期刊的借阅权限请求及还书请求数据。服务响应数据指图书馆内管理人员即服务提供者对服务接受者的请求做出的回应数据,如电子显示屏上显示的借阅期限或借阅、还书成功的提示语句。

对图书馆内的数据进行产生对象的分类有助于完善图书馆服务,收集分析反馈信息,对服务质量进行有效调整,提供智能化服务。

3.5 按照人工处理参与度

根据人工参与的程度可将数据分为不同等级。根

chinaXiv:202304.00374v1

李玉海，张驰，王文卓.高校图书馆数据流失与保全对策［J］.图书情报工作,2020,64(1):139－145.

据这些不同的等级可将数据进行分类。例如人工参与度高的数据有部分管理型数据、图书借阅信息等与图书馆流动人员有关的数据,人工参与度低的数据有图书索书号、书籍上的文字等客观存在的数据。

在图书馆智能化管理的进程中,呈现人工参与度逐渐降低的趋势,如古代图书馆图书资源少,管理封闭;近代图书馆更加普及和开放;现代图书馆引入信息技术和网络技术,管理效率大大提升[2-3]。

3.6 按照数据载体

以载体为标准可将高校图书馆数据分为设备数据、录入数据、纸本数据和电子数据。

设备数据是指自动感应记录的数据,例如图书馆一天内人员进出的流量。录入数据指人工录入的数据,如某一学生的一学年的图书借阅信息。纸本数据指图书馆内的纸质图书期刊等。电子数据可体现在视频、音频、文本和图形等。随着信息技术的发展,电子数据日趋多样性,但电子数据的有效利用程度仍然不高,对纸质数据和电子数据的使用情况进行长期分析有利于电子资源的普及。

4 数据保全必要性

数据保全即对数据进行整理、加密,使其不随时间变化并保存起来以供后续分析使用。高校图书馆的数据具有保全价值。基于数据特点,管理服务型数据的保全有利于在长时间段内的分析和整合,优化图书馆服务,打造智慧图书馆;数字资源的保藏有利于后期借阅和资源价值的反复利用;图书资料型数据需进行系统机构型的更新维护,以提供给大学生更好的学术性服务。

虽然国外对数据保全的话题逐年递增(见表1),但是国内却很少有人能够意识到数据保全的重要性以及重大意义,数据保全可以为图书馆决策(如购书策略、发展方向、数据追溯研究等)和发展提供重要的科学依据。另外,数据保全对未来回顾如今发展现状以及过程从而提出新的发展方向、追溯最早的数据来源等方面具有极其重要的研究价值。数据保全对图书馆现有的价值及其必要性体现在以下几个维度。

4.1 档案性

有些数据具有档案价值,必须保全才能发挥作用。档案价值是指一些数据具有存储整理的必要性,以供后期查阅。例如图书馆的馆藏信息具有档案价值。当图书馆管理人员发生流动时,新晋管理员可根据保藏下来的信息核对图书馆的书籍期刊文献数量，进行日常图书馆信息的维护和更新。从管理层面来看,对于图书馆相关人员的信息也需留存档案,以备后期责任追究和人员调动。

表1 数据保存文献数增长情况(来源:爱思唯尔数据库)

Key Words: Data Preservation

年份	篇数
2019	23 085
2018	20 009
2017	18 134
2016	17 715
2015	16 524
2014	15 235
2013	14 332
2012	13 235
2011	12 161
2010	11 061
2009	10 843
2008	10 571
2007	9 900
2006	8 977

4.2 过程管理

数据是重大事件发展的线索,是图书馆管理过程的体现。图书馆数据具有动态性可以反映图书馆管理进程,任一时刻产生的数据都有可能具有后期查阅的价值。如软件开发中的版本控制,在软件迭代的过程中,可能某一版本具有不可替代的优越性,版本数据的留存为后期版本更新提供参考。又如规章制度的修订拟定,图书馆根据每一阶段的管理进程进行合理更改,每一年出台的规章制度的细微之处的修订都应妥善留存。

4.3 工作效率

数据具有时效性。数据的妥善留存为任务继承做准备,可以大大提高工作效率。进程数据的保存可为下一阶段工作节省工作量,避免重复冗余的操作。如图书馆数据的录入、分类和管理有助于管理人员进行查找、搜寻而后利用。相关文件的保存备份可大大减少工作量,利于相关工作人员在此基础上进行更改和完善。

4.4 个性化服务

对用户行为数据保全得越全,个性化服务就越有针对性。相关数据需要进行合理的整合分类。对于图书用户的个人信息需妥善加密,相关行为的分析利于后续个性化服务的定制。

在采集用户数据并保藏的过程中,需要注意时间连贯性、数据采样密度、样本数量、数据精度等。首先,

chinaXiv:202304.00874v1

ChinaXiv合作期刊

时间连贯性基于数据的时效规律性。用户某一段时间的行为数据相互之间含有隐藏信息,忽略其中一段数据的储存会对分析的精确性造成影响。其次,数据采样密度和样本数量是指用户人群的采样数量,对于一类相似行为数据,采样数量应合理,有时候太精确无意义。例如,系统分析大学新生借阅图书情况时,需要连续采集新生入学后几个月的借书情况,在前两个月用户数据较少,可能与军训和适应期有关,采样密集度、样本数量相对较低。最后,数据精度应基于实际样本对象和问题具体分析。

5 数据流失问题

从数据流失的形式和原因,以及国内外高校图书馆数据流失和保全现状展开讨论,可以让读者对国内外数据流失的情况以及具体环节有一个基本的认识,从而为改善并解决数据流失问题做准备。

5.1 数据流失形式及成因

管理过程流失及原因。管理型数据主要针对图书馆管理人员,往往由于管理过程中的失误而导致数据失去其价值或未妥善保存。

在领导层面上,由于管理人员的疏忽和不完善的反馈机制,导致需要引起注意的数据被忽视,如某一时间段内的图书借阅量突然增加,而未分析其原因,或某一数据库的需求量增加而未引起重视;在组织层面上,内部人员分工不当导致关键人员离职或岗位的缺失会导致图书馆管理上的混乱[4],如书籍的分批采购、数据库的更新等,需各尽其职;在计划层面,高层管理人员缺乏数据敏感性,对关键信息的提取缺乏经验,在高校图书馆的环境下缺乏针对性管理措施;在控制层面,相关管理人员缺乏预测能力,可通过软件计算解决。

技术性流失及原因。技术性数据丢失是客观的,由相关设备不完善导致。在数据收集上,如下载量数据的更新滞后、灵敏度和精度不高,会对数据分析造成影响。在数据组织上,高校图书馆数据的专业性要求强,缺乏相应对策和标准,会使文献资源数据失去其本身价值;在数据传输上,通信渠道缺乏可靠性也是数据流失的重要原因[5];在数据存储上,未根据数据类型进行合理分类,使具有长期存储价值的数据无有效存储空间,或过度信赖存储机制,导致偶然的数据丢失无法恢复,在数据的备份管理上未加强有效监管。

自然性流失及原因。自然性流失包括潜在需求数据流失、客观规律数据流失及关联业务数据流失。数据之间具有一定的关联性。在技术不精、管理不完善的条件下,某些重要隐形数据未被关注。同时数据具有时间规律性,如新生入馆期间的数据与以往其他时段是不同的并存在潜在变化,数据未被发掘导致管理层面的疏忽。与图书馆相关联的数据还有网络中心相关数据、财务资产数据、运营管理数据等,将图书馆数据孤立分析,导致相关有效数据未得到重视。

5.2 国内外高校图书馆数据流失情况

国内高校图书馆数据流失情况。笔者通过对清华大学、北京大学、武汉大学、华中科技大学、上海交通大学等高校官网进行分析,以研究我国双一流大学图书馆数据流失以及保全的现状,发现北京大学、上海交通大学的图书馆官网设计较有特色且具有一定的数据保全意识,而上海交通大学在图书馆大数据方面表现得更加出色。上海交通大学图书馆自 2009 年以来,每年都有年度报告,其中上海交通大学图书馆 2018 年度报告[6]首页包含了到馆人数、开放时长、总借还量、培训次数、论文发表数、服务人次等众多数据,并进行可视化展示,在数据开放获取方面意识较强。另外,华中科技大学链接其特色的创业服务,北京大学其研究生 + 教师的特色服务说明其对学生的数据服务已有所起色。但其他大多数高校都没有做到对自身数据进行可视化分析,且大部分图书馆都只提供第三方数据库的目录以及链接,其中链接到自身高校数据库的少之又少,国内高校图书馆要做到保全自身数据且灵活运用仍需较长一段时间。高校图书馆的数据保全不能局限于图书馆自身,作为大学的重要组成部分,高校图书馆还需与学校其他部门和组织进行深入合作,为大学的文化传播、教师水平、学生综合素质的提高提供全面的帮助[7-8]。

国外高校图书馆数据流失情况。通过对国外剑桥大学、斯坦福大学、哈佛大学、麻省理工学院等高校的图书馆官网进行分析,剑桥大学已正式将数据保全作为其发展的第二高风险,并提出了剑桥大学数字保存战略规划[9]。自 2016 年,牛津大学与剑桥大学开展了名为牛津和剑桥的数字保存计划(Digital Preservation at Oxford and Cambridge Project),其主要目的为在大学内部创建知识和研究方面的数字馆藏,并影响到国际研究界以及更多的社会组织。剑桥、哈佛等图书馆与其自身高校文化结合的较好,为其师生提供了该校具有较高研究价值的特色资源,从其官网的建设中可以体现出其文化氛围及学习气氛。虽然其数据开放利用程度较高,但在其战略规划中指出用 DPCMM 模型来分析其数字保存情况,图书馆的数字保存计划仍然还

chinaXiv:202304.00034v1

李玉海，张驰，王文卓.高校图书馆数据流失与保全对策［J］.图书情报工作,2020,64(1):139－145.

chinaXiv:202304.00374v1

有很长的路要走。

虽然国内有少数高校已逐渐开始制定图书馆数据保全政策,但是无论国内还是国外,数据流失的状况都在发生,高校图书馆不同于公共图书馆,高校图书馆想要带给学生更好的数据服务,与高校其他部门和组织的合作势在必行。而图书馆为自身发展与改善所需收集并保存的数据也有很多需要研究的地方,对比国内外高校图书馆的数据流失和保全情况,虽然调查中的国外高校图书馆做的好一些,但是差距并非很明显,仍有赶超的机会。

6 数据保全策略

数据保全在司法范畴应用较多,在图书馆应用中应合理借鉴其特性,考虑图书馆数据的特点,以及相关设备、工具的支持,对保全的效力、范围进行分类精确考量。下面从图书馆业务、管理、创新、方法和工具几个层面展开讨论,另外将在结尾处借用 DPCMM 模型为图书馆的数字保存策略提供思考点。

6.1 基于基本业务

数据保全应基于基本业务开展的过程,开展业务时需确定数据类型、明确管理岗位、梳理工作规程,然后根据在此过程中产生的数据设置数据保全时间、保全精度和保存价值。例如在进行图书馆日常书籍期刊的维护工作时,根据书籍期刊的类型、借阅和借出的时间、借阅次数、使用时间段进行分类。图书借阅时间的保全有助于管理人员分析时效规律性,为后续管理提供参考。借阅时间的保存有利于提高图书管理员的工作效率,避免不必要操作。又如在进行工作年报、事实数据申报及考核等工作时,需区分文件类型,合理保藏档案,对于每递交上的文件进行整理和归类。

6.2 基于管理效率

基于管理效率考虑图书馆保藏策略时,时间是重要的参数。对管理过程中的数据按照时间进行划分,有利于提高管理效率。如数据采集数据、响应数据、报表生成时间等。数据采集数据是指数据的采集性质的设定,在数据采样过程中,不是全部数据都需保藏,应根据时间点划分,留下关键性数据,同时也利于节省存储空间[10]。响应数据是指相互关联的数据的响应。有些数据响应快,有些数据需经过一段时间才可获得。对于响应数据的存储应注意根据响应时间结点进行保藏。报表生成时间与数据生成效率有关,数据报表需保证数据来源的可靠性,并根据数据复杂度确定其生成时间。

6.3 基于管理成本

根据管理成本对数据保藏策略进行综合考量,可确定数据保全的程度。从人力成本来看,图书馆人工参与度越来越低,需对管理人员操作的数据和机器产生的数据进行分类,进而细化人力成本等级。一般而言,人力成本越高,数据越复杂,存储价值越高。从财力成本来看,部分数据耗费大量财力资源,对数据的保藏应更加精细。如外文外刊书籍的收集需耗费大量人力物力,对其维护和监管需更加细致。从物质成本来看,部分数据基于物质产生,如相关设备的耗材、系统的定期维护等,基于物质成本划分也是对数据价值的一种考量。

根据相关标准对数据进行分类后,数据的保藏需考虑备份时间、冗余设备配置、安全权限等级等[11]。备份时间是指对数据的保存时间,一般以月或年为单位,对数据进行更新。冗余设备配制需考虑设备的工作效率,如关键岗位的 UPS 配制是否可支持当前数据存储量、U 盘移动硬盘是否安全可靠等,需定期进行检查更新。

6.4 基于创新管理

面对图书馆的创新服务,可分为如个性化服务、泛在服务、智慧图书馆等。个性化服务是指对图书馆使用者进行个性化推荐、智能还书提醒、相关馆藏书籍定位信息推送等,是通过用户行为数据分析得到的[12]。泛在服务是指扩大服务范围,打造常态性服务。泛在服务的数据来源于不同基层人员的需求和相应的服务策略。智慧图书馆是指利用互联网数据打造智能化管理的图书馆。

基于创新管理的图书馆数据保藏策略可根据相应管理手段来考量。例如对图书馆数据的全生命周期进行技术改造、发明创造,需要根据时间结点收集用户数据,得到实时用户数据反馈,以及调整服务方式。如建立基于图书借阅数据的柔性书库、基于用户行为数据的个性化推荐、基于用户行为数据的文献资源配置等,需根据特定用户某一时段借阅数据,根据用户人群进行划分,得到准确的有针对性的数据,在一定时效内进行数据的积累和分析。

6.5 数据保全方法工具

6.5.1 数据保全的方法

将数据进行分类后,基于不同情景下运用的保藏策略不同,相应的保藏方法也有差别,合理运用方法有利于提升工作效率[13]。首先,通过程序设计可以解决日常办公中的事件捕获问题。如抓取图书期刊号前四

ChinaXiv合作期刊

位或整合一天内图书借阅类别和期刊信息。其次，对生产、管理和服务中产生的数据需进行合理截取和整理。对日常办公数据应统一版本，方便协同工作。并对全部数据进行备份，对某些数据设置保藏时间以节省存储空间。其中图书馆需保存的主要方法及场景如表 2 所示：

表 2 数据保存方法及场景

方法	应用场景
程序设计	事件捕获，cookies 日志文件
台帐收割	对生产、管理和服务数据
版本控制	对日常办公数据
备份	全部数据

6.5.2 数据保全的工具

数据保全可利用的工具有程序类、平台类和存储类。对于数据的获取，常用工具有火车头、八爪鱼、猪八戒等网络爬虫工具。管理人员可对相关图书信息库进行爬取，获取关键信息，查找相关图书编号、借出借阅信息、相关图书分类信息等。还可建立专业管理平台，便于图书馆使用者查询图书在馆信息、座位使用信息及自习室借用信息等。图书管理员可根据借阅者过往使用信息及情况快速定位，提高工作效率。还可建立 Raid 桌面存取、云存储平台进行数据的存放和查找，延长数据保藏时间。

6.6 数字保存能力成熟度模型(DPCMM)

数字保存能力成熟度模型(DPCMM)于 2007 年首次被威斯康星州大学的 L. Ashley 教授和加拿大英属哥伦比亚大学的 C. Dollar 教授提出[14]。DPCMM 将机构数字保存能力分为最低、较低、中级、高级、最优 5 个阶段，其中有 15 个关键审核标准[15-16]。所有 15 个 DPCMM 都符合 ISO14721 的要求和 ISO16363 中的审核标准组件要求，才能达到中级(第 3 阶段功能)水平，见图 2。

(1)数字保存基础设施：包括政策、战略、管理、合作、技术知识、开放标准、特定社区、电子文件调查。

(2)数字保存服务：包括摄取、存储、设备/媒体更新、完整性、安全性、保存元数据、访问。

DPCMM 在详细描述每个审核标准的概念后，给出 5 个层级(分别对应 0-4 分)对 15 个审核标准进行打分，从而将机构数字保存能力进行评级。其量化的指标及概念对国内高校图书馆的数据保全有着指导性作用。

7 结语

图书馆是大数据的集合地。大数据是短期内难以

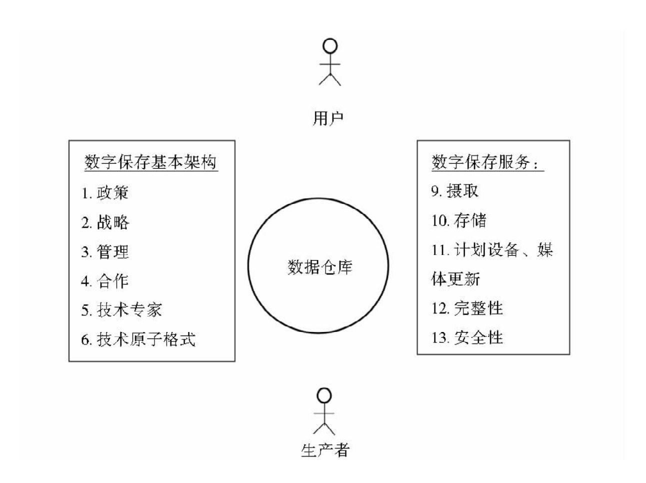


图 2 数字保存能力成熟度模型框架

形成的数据，靠长时间的积累[17]。高校图书馆数据除具有数据的基本特点外，还具有基于高校背景下的特有特点。由于高校图书馆主要的服务对象为高校大学生，不同于公共图书馆的服务对象及馆藏数据的广泛性，高校图书馆数据的存储和管理具有时效规律性、专业性，往往随着开学季毕业季呈现出波动性和相对稳定性。

图书馆的数据是有规律、有价值、有意义的，也必然会形成大数据。高校图书馆数据的流失主要体现在相关管理数据未得到重视，或与用户相关的行为数据未得到正确分析[18]。如学生考试周以及考研等数据集中的阶段，未对数据进行合理保藏和利用等。

现在由于认识和意识的局限，客观存在数据流失现象。随着图书馆创新管理工作的进行，以及智慧化图书馆建设和利用的跟进，在高校图书馆走向智能化的背景下，数据的保藏将越来越受到重视，数据流失问题也终将被解决。

参考文献：

[1] 黄宗忠. 论图书馆核心价值(下)[J]. 图书馆论坛,2008(1):1-3.
[2] 张金梅. 数字环境下公共图书馆资料长期保存策略研究[J]. 才智,2019 (2):299.
[3] 乔文明. 网络环境下的高校图书馆的信息资源建设[J]. 图书馆论坛,2001 (5):59-61.
[4] 郑勇. 话语视角下高校图书馆馆藏资源利用现状与对策[J]图书馆建设,2009(4):115-120.
[5] 李成. 大数据时代高校图书馆数据监护要素与优化策略研究[J]. 科学管理,2019 (1):58-64.
[6] 上海交通大学图书馆 2018 年度报告[EB/OL]. [2019-12-01] http://www.lib.sjtu.edu.cn/sjtu/upload/ueditor/file/20190605/Library%20Annual%20Report%202018.pdf.
[7] SHIMRAY S, RAMAIAH C. Digital preservation strategies: An overview. (2018)[EB/OL]. [2019-12-01]. https://www.re-

chinaXiv:202304.00034v1

searchgate. net/publication/327221006_Digital_Preservation_Strategies_An_Overview.
［8］刘斌,黄婧,李峰,等. 大数据环境下高校图书馆数据管理与分析应用实践——以北京师范大学图书馆为例［J］. 数字图书馆论坛,2018(10):54－59.
［9］LANGLEY S. Cambridge University Libraries digital preservation strategy［EB/OL］.［2019－12－01］. https://www. repository. cam. ac. uk/bitstream/handle/1810/291864/STR_CUL_digitalpreservationstrategy_v1_20190415. pdf? sequence＝5.
［10］储节旺,张瑜,刘青青. 高校图书馆“双一流”建设的战略思考［J］. 大学图书馆学报,2019,37(1):6－16.
［11］BOTE J, FERNANDEZ-FEIJOO B, RUIZ S. The cost of digital preservation: a methodological analysis［J］. Procedia technology, 2012,(5): 103－111.
［12］王晓冉. 高校图书馆技术竞争情报服务模式研究［D］. 南京：南京航天航空大学,2015.
［13］李凌杰. 从学科建设角度优化馆藏文献信息资源配制研究——以天津科技大学为例［J］. 图书馆工作与研究,2009(6):50－53.
［14］孙超,吴振新. 国外数字资源长期保存成熟度模型及其分析与评价［J］. 图书情报工作, 2017,61(1):32－39.
［15］DPCMM white paper revised April 2014［EB/OL］.［2019－11－26］. http://www. securelyrooted. com/dpcmm.
［16］PEARSON D,COUFAL L. Digital preservation environment maturity matrix［R/OL］.［2019－11－29］. http://www. nsla. org. au/sites/www. nsla. org. au/files/publications/NSLA. Dig Pres_Environment_Maturity_Matrix. pdf .
［17］马晓亭. 大数据时代图书馆数据可用性:价值、挑战和保障［J］. 图书馆理论与实践,2014(10):5－8.
［18］程晓岚,马海群. 国内外高校图书馆馆藏采购政策比较研究［J］. 2016(9):73－79.

作者贡献说明：
李玉海:提出论文选题与研究思路,提出论文框架;
张驰:数据及资料采集与处理,撰写并修改论文;
王文卓:资料分析,撰写论文。

chinaXiv:202304.00374v1

Data Loss and Preservation Countermeasures of University Library

Li Yuhai[1,2]　Zhang Chi[1]　Wang Wenzhuo[1]
[1] School of Information Management, Central China Normal University, Wuhan 430079
[2] Research Center for Innovation and Development of Chinese Libraries, Central China Normal University, Wuhan 430079

Abstract: **［Purpose/significance］** To investigate the data preservation capability and loss status of some double-first class university libraries in China, research and analyze the application of the maturity model of digital preservation capability in reducing data loss and data preservation, and for the construction and development of domestic data preservation systems Provide references. **［Method/process］** By exploring the data preservation and utilization of the official websites of some university libraries at home and abroad, it analyzed from the dimensions of digital collection resources, digital preservation capabilities and policies, and digital preservation consciousness. Comments on the resulting data loss. **［Result/conclusion］** To summarize the current status of digital preservation in some colleges and universities in China, and put forward the necessity and significance of improving the digital preservation capabilities of Chinese university libraries, and use the Digital Preservation Capability Maturity Model (DPCMM) to provide theoretical support for university library data preservation.

Keywords: university library　data loss　preservation